

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-175385

(43)Date of publication of application : 30.06.1998

(51)Int.Cl.

B42D 15/02

G06K 9/03

(21)Application number : 08-353877

(71)Applicant : DAINIPPON PRINTING CO LTD

(22)Date of filing : 19.12.1996

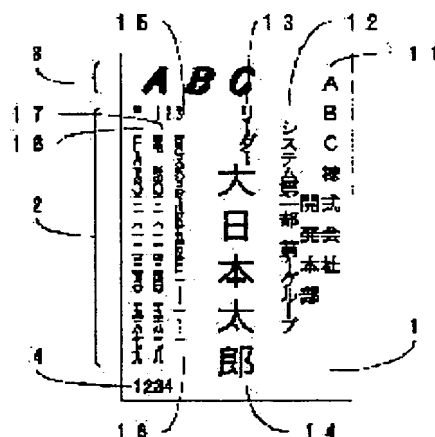
(72)Inventor : NAKAYASU EIJI

(54) PRINTED MATTER WITH INSPECTION CHARACTER

(57)Abstract:

PROBLEM TO BE SOLVED: To inspect mechanically whether the erroneous recognition is or not on a character string read by a character recognition device and save the labor for verifying a character string data by providing inspection characters for a character string of the Japanese language including Chinese characters to be read mechanically.

SOLUTION: When character figures formed by characters printed on a namecard such as the name of organization 11, a department signature 12, a title 13, a name 14, an address 15 and a telephone number 17 are converted to a character data, proper codes are provided for characters constituting respective items, and character codes are handled as symbols representing the kinds of characters in a computer. In the above-referred name card assembly, the sum total of character codes for all mechanically read characters except log marks 3 is found and then four character-inspection characters 4 [1234] are obtained in the given order and printed on margins of the namecard.



LEGAL STATUS

[Date of request for examination]

17.07.2003

[Date of sending the examiner's decision of rejection]

20.06.2006

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-175385

(43) 公開日 平成10年(1998) 6月30日

(51) Int.Cl.⁶

B 4 2 D 15/02

G 0 6 K 9/03

識別記号

5 2 1

F I

B 4 2 D 15/02

G 0 6 K 9/03

5 2 1

D

審査請求 未請求 請求項の数 4 F D (全 7 頁)

(21) 出願番号 特願平8-353877

(22) 出願日 平成 8 年(1996)12月19日

(71) 出願人 000002897

大日本印刷株式会社

東京都新宿区市谷加賀町一丁目 1 番 1 号

(72) 発明者 中保 英司

東京都新宿区市谷加賀町一丁目 1 番 1 号

大日本印刷株式会社内

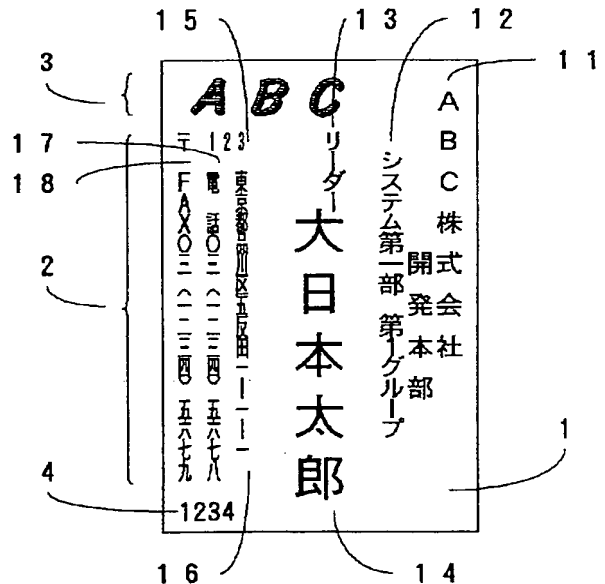
(74) 代理人 弁理士 小西 淳美

(54) 【発明の名称】 検査文字付き印刷物

(57) 【要約】

【課題】 印刷物に印刷されている日本語文字を認識してデータ処理装置に入力する際、読み取った文字列の誤認識の有無を簡便かつ正確に検査することができる印刷物を提供する。

【解決手段】 漢字を含む日本語の機械読取対象文字列 2 に対して、検査文字 4 を付した検査文字付き印刷物であり、前記印刷物は、名刺 1 であり、さらに、前記印刷物の検査文字 4 が、10進数の数字で構成されているか、または、16進数の英数字で構成されている検査文字付き印刷物である。



【特許請求の範囲】

【請求項1】 漢字を含む日本語の機械読取対象文字列に対して、検査文字を付したことを特徴とする検査文字付き印刷物。

【請求項2】 前記印刷物は名刺であることを特徴とする請求項1に記載の検査文字付き印刷物。

【請求項3】 前記検査文字が10進数の数字で構成されていることを特徴とする請求項1または請求項2記載の検査文字付き印刷物。

【請求項4】 前記検査文字が16進数の英数字で構成されていることを特徴とする請求項1または請求項2記載の検査文字付き印刷物。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、漢字を含む日本語の文字列を機械読取する印刷物に関するものである。さらに詳しく述べるならば、漢字、ひらがな、カタカナ、英数字、記号などで構成された文字列を、機械読取するときに、認識文字の誤りの有無を判別するための検査文字を付した印刷物に関するものである。

【0002】

【従来の技術】近年、画像読取装置の光電変換素子の精度が向上し、また文字認識ソフトの文字認識率が向上し、安価な文字読取装置が一般化しつつある。例えば名刺専用の文字読取装置や、新聞記事やワープロで作成した文書を、パーソナルコンピュータや携帯情報端末機にデジタルデータとして入力できるテキストリーダーと称する文字読取装置がある。

【0003】例えば、名刺専用の文字読取装置では、名刺に印刷されている文字列を光学走査し、会社名、所属、肩書、氏名、及び郵便番号、住所、電話番号、ファックス番号などの文字画像を認識処理して、デジタル化された文字列データに変換し、パソコンやワープロなどの情報機器に名刺情報として伝送して、住所録や顧客管理、ダイレクトメールの宛て名印刷などに利用することができる。また、テキストリーダーでは、漢字を含む日本語の文章を印刷した印刷物を、光学走査し、文書の読取領域を自動判別して、文字認識のうえ、読み取った文字列データをパソコンなどに入力し、文書情報のデータとして使用することができる。

【0004】

【発明が解決しようとする課題】しかし、これらの名刺読取装置やテキストリーダーなどの文字読取装置における文字認識能力は、以前に比べると格段に上昇したが、読み取る印刷物の汚れや、印刷した文字のかすれなどの影響で、まれに誤認識が発生することが避けられない。特に漢字を含む日本語の文字列では、例えば、「大」と「太」と「犬」、「月」と「目」、「日」と「田」のように非常に良く似た字形の文字が含まれている場合は、誤認識の可能性がある。従って、現状では、文字読取装

置で名刺やワープロ文書、あるいは新聞記事などをデジタルデータ化した場合、文字認識に誤りがないかどうかを検査する必要がある。

【0005】この検査はすべての文字、一字一句について行なわなければならない、長文の文章では大変な労力を必要とし、また、仮に検査を行なったとしても、すべて正しいという保証はなく、現実的な方法とはいえない。本発明は、上記文字読取装置における文字認識において、読取文字の認識ミスを機械的に自動検出するのに適した印刷物を提供しようとするものである。

【0006】

【課題を解決するための手段】本発明は、名刺やワープロ文書、その他の日本語印刷文書において、機械読取用の検査文字をあらかじめ印刷しておき、文字読取装置にて、読み取った文字の認識ミスを検査する機能を付加しておくことにより、機械読取対象文字列の認識ミスの有無を自動的に検査する。

【0007】すなわち、本発明の印刷物は、請求項1に示したように、漢字を含む日本語の機械読取対象文字列に対して、検査文字を付した検査文字付き印刷物であり、前記印刷物は、請求項2に示したように、名刺であり、さらに、前記印刷物の検査文字が、請求項3に示したように10進数の数字で構成されているか、または、請求項4に示したように16進数の英数字で構成されている検査文字付き印刷物である。

【0008】

【発明の実施の形態】以下本発明の実施例に基づいて、発明の詳細を説明する。図1は、本発明の検査文字付き印刷物の一例である名刺の平面図である。図において、11は、名刺所有者の所属する団体名であり、12は、前記団体の組織の一部である所属部署名である、13は、名刺所有者の役職名、または資格であり、14が名刺所有者の姓名である。また14～18は名刺所有者の連絡先である住所や電話番号などである。これらの項目は従来の名刺と何ら変わりがなく、従来どおりの書式で印刷されている。

【0009】図1の実施例に示す名刺には、前記所属団体名や所属部署名、本人氏名や役職名、住所や電話番号とともに、名刺の最下端に4桁の数字が印刷してあり、この数字が本発明でいうところの検査文字4である。前記4桁の数字で印刷されている検査文字4は、例えば、以下に説明するような方法で計算された数値を元に印刷する。

【0010】すなわち、前記団体名11や所属部署名12、肩書13や姓名14、所在地15、電話番号17など名刺に印刷されている文字は、前述した、名刺読取機やテキストリーダーなどの文字読取装置において、文字図形から文字データに変換されたとき、前記項目を構成する文字にはそれぞれ固有のコードが与えられ、その文字コードがコンピュータの中で文字の種類を表す符号と

して扱われる。

【0011】上記文字コードには各種のコード体系があり、コンピュータメーカや機種によって異なったコード体系の文字コードが使用されている。一方、メーカや機種によってコード体系が異なるとデータの互換性がなくなるので、JISによって定められた共通の文字コードとしてJIS X 0208がある。ちなみに、数字の「0」～「9」および英字の「A」～「F」の文字コードを、区点コード、およびJISコードで表現すると、下表に示すような文字コードが付与されている。

【0012】

文字	区点コード	JISコード
0	0316	2330
1	0317	2331
2	0318	2332
3	0319	2333
4	0320	2334
5	0321	2335
6	0322	2336
7	0323	2337
8	0324	2338
9	0325	2339
A	0333	2341
B	0334	2342
C	0335	2343
D	0336	2344
E	0337	2345
F	0338	2346

【0013】上表に示すように、各文字に付与される区点コードは、左2桁の「区番号」を「01」～「94」の10進数で表し、右2桁の「点番号」を「01」～「94」の10進数で表す。同様にJISコードは、2バイト16ビットの2進符号を、16進数の英数字で表現しており、左2文字の上位バイトが「21」～「7E」、右2文字の下位バイトが「21」～「7E」の値をとる。すなわち、コンピュータの内部で、日本語の文字は、上記JISコードあるいはシフトJISコードなどで定義されているコード体系に基づいて、1文字を2バイトのデータで表して記憶されている。

【0014】次に、検査文字の計算方法について説明する。例えば「ABC株式会社」と印刷される文字列に対して、検査文字を計算する場合、まず、対象となる文字の文字コードが、次のように与えられているものとする。なお、下記において、各文字に与えられる文字コードは、JIS X 0208に定義されている区点コードで表現している。

「A」=0333

「B」=0334

「C」=0335

「株」=1984

「式」=2816

「会」=1881

「社」=2850

【0015】次に、上記で与えられた、各文字コードの総和を求める、上記の例では総和=10533となる。次いで、前記総和の下4桁「0533」を得る。次に4桁の数値を10000から引き「9467」を得る。このようにして得られた4桁の数値を数字に変換する。このようにして得られた4桁の数字「9467」が、「ABC株式会社」に対する検査文字となる。

【0016】なお、図1に示した名刺の例では、ロゴマーク3を除く全ての機械読取対象文字に対して、付与されている文字コードに基づいて、その総和を求めた後、前記した手順で4文字の検査文字「1234」を得、これを、名刺の余白部に印刷している。

【0017】次に、図1に示した名刺の文字を読み取る文字読取装置について説明する。図3は、本発明の印刷物を読み取る文字読取装置のブロック図である。図において、5は、文字認識装置であり、6は、データ処理装置である。

【0018】文字認識装置5は、図に示すように、画像読取部21、文字認識部22、文字列編集部23、およびデータ伝送部24とで構成されている。画像読取部21は、光電変換素子で構成され、1次元のラインセンサや2次元のエリアセンサを用いて、印刷物上の印刷情報を白黒の2値画像データに変換する。文字認識部22は、前記2値画像データに含まれる文字画像を順次抽出して、文字形状データと照合し、文字画像に最も近い形状の文字を選択決定する。また、文字列編集部23は、文字認識部22で選択決定した文字列データに基づいて、各文字列を、図2に示すように、文字列の属性ごとに分離して仕分ける。データ伝送部23は、画像読取部21で読み取った印刷物の2値画像データと、文字認識部22および文字列編集部23を介して認識編集した文字列データを、データ処理装置6に伝送する。

【0019】なお、図示していないが、文字認識装置5には、画像データや文字列データを一時記憶するためのメモリや、文字認識装置5を構成する各部の動作を制御するためのマイクロコンピュータが内蔵されている。

【0020】データ処理装置6は、データ受信部31、文字検査部32、文字修正部33、およびデータ記憶部34で構成される。データ処理装置6は、例えば、パーソナルコンピュータで構成され、パーソナルコンピュータが備えているハードウェア資源、およびソフトウェアを利用して、前記各部を構成することができる。例えば、データ受信部31は、パーソナルコンピュータの通信装置とそれを制御する制御プログラムによって構成す

る。また文字修正部33は、パーソナルコンピュータに接続されたディスプレイに修正すべき文字列とその画像データを表示し、画像データを目視で確認したうえで、キーボードを用いて文字列データを修正するように構成する。また、データ記憶部34は、パーソナルコンピュータに備えている固定ディスク装置やフレキシブルディスク装置で構成する。なお、文字検査部32は、ソフトウェアのプログラムで簡単に構成することができる。

【0021】例えば、前記4文字の検査文字を使用して文字の認識間違いを検査する方法は、次の手順を実行するプログラムを構成することで実現できる。前述したように、「ABC株式会社」に対する文字コードの総和は「10533」であり、検査文字は「9467」となっている。従って、まず、文字列に対する文字コードの総和を求める。もし文字列を正しく文字認識したときは、その総和は「10533」になる。次に、検査文字の文字列から検査値を求める。次いで、前記総和と検査値との総計を求める。例の場合、総計は $10533+9467=20000$ となる。文字列検査は、前記総計の下4桁が「0000」になるかどうかで判断する。もし、総計の下4桁が「0000」になれば、文字認識が正しく行われたと判断し、総計の下4桁が「0000」以外のときは、文字認識ミスがあったと判断する。

【0022】例えば、前記文字列において「B」を、「8」の文字に誤認識した場合、「8」の文字コードは「0324」なので、「ABC株式会社」に対する文字コードの総和は10523となり、その結果、検査文字を含めた文字コードの総計は「19990」となり、総計の下4桁が「9990」になる。従って、総計の下4桁が「0000」以外なので、前記文字列のいずれかの文字を誤認識したことが判明する。以上の手順をプログラミングすることによって文字列検査部32を構成する。

【0023】次に上記文字読取装置の機能動作について説明する。図5は、本発明の検査文字付き印刷物である名刺の文字読取装置の一例を示す斜視図である。図において、5は、文字認識装置であり、文字認識装置5には、名刺を複数枚搭載して、1枚ずつ画像読取部に供給するためのホッパ51を備えている。ホッパ51には、名刺を20~100枚程度、束ねて搭載することができる。読取可能な名刺の寸法は、幅48~55mm、長さ85~91mmで、紙厚0.15~0.45mm程度である。

【0024】図示していないが、文字認識装置5の内部には、名刺画像を読み取るための光電変換素子を備えており、例えば、約16画素/mm(400ドット/インチ)の画素密度を有するリニアイメージセンサを備えている。前記画素密度の光電変換素子を用いることによって、一般的に名刺に印刷されている2~10mm程度の大きさの文字を高分解能で読み取ることができる。

【0025】また、図示していないが、文字認識装置5の内部には、前記光電変換素子で読み取った名刺画像の印刷文字を文字認識するための文字認識部が内蔵されており、読取可能な文字種として、JIS第1水準の漢字2965字、およびJIS第2水準の漢字3388字のうち、名刺に使われる頻度の高い文字約600字、さらに数字10字、英字52字、ひらがな73字、カタカナ83字、その他記号約150字など合計4000字程度の文字図形辞書を備えている。

【0026】さらに、図示していないが、文字認識装置5の内部には、文字認識装置で認識した文字列データを、外部に接続されたデータ処理装置に伝送するためのデータ伝送装置を内蔵している。このデータ伝送装置は文字列データをシリアルデータに変換して伝送する機能を備えており、パーソナルコンピュータなどと簡単に接続して読取データを出力することができる。このデータ伝送装置は光電変換素子で読み取った名刺画像データ、および文字認識の結果の文字列データを、信号線56を介して、文字読取装置5から、データ処理装置6に入力する。

【0027】データ処理装置6は、いわゆるパーソナルコンピュータであり、文字認識装置5で読み取った名刺画像や文字認識の結果を表示するディスプレイ62と、文字認識結果を必要に応じて修正するためのキーボード62を備えている。また図示していないが、読み取った名刺画像や文字認識の結果、または修正後の文字列データを記憶しておく記憶装置を内蔵している。

【0028】さらに、前記記憶装置には、文字認識装置5で読み取った画像データと、前記画像データに含まれる文字画像の認識結果を利用して、住所録や顧客名簿を作成するためのデータ処理プログラムが内蔵されている。また、通信線56を介して、文字認識装置5の動作を制御する制御プログラムを前記記憶装置に内蔵しておき、パーソナルコンピュータを使用して、文字認識装置5を遠隔操作することもできる。

【0029】上記構成の文字読取装置で、図1に示すような名刺の印刷文字を読み取る方法について説明する。まず、文字認識装置5のホッパ51に、読み取る名刺の束を向きをそろえて装填する。次に、データ処理装置6において、文字認識装置5用の制御プログラムを起動し、文字認識動作を開始させる。文字認識装置5ではホッパの名刺束から、1枚を抜き出して文字認識処理に入る。

【0030】文字認識処理は、図3に示した画像読取部21と文字認識部22において行われる処理である。まず、画像読取部21の光電変換素子によって名刺の表面が光学走査され、印刷画像が2値の画像データに変換されて、一旦メモリに蓄えられる。次に、メモリに蓄えられた画像データのなかから、文字認識部22が文字画像を順次抽出し、次いで、あらかじめ登録されている文字

図形辞書と照合して、抽出した文字画像に最も近い形状の文字図形を選択し、文字種を決定する。次に、すべての文字画像について文字種を決定したら、文字列編集部23によって文字列の整理を行う。

【0031】図2は、図1に示した名刺を文字認識装置で処理した結果を示している。図2に示すように、文字認識された文字列は、各文字列に含まれる特定の文字や記号、または、文字列の配置されていた位置によってその属性が決定される。例えば文字列のなかに「株式会社」または「有限会社」などの法人を表す文字が含まれているとき、その文字列は名刺所有者が所属する法人団体の名称であると判断できる。また、文字列の中に都道府県名や市区町村名と番地に該当する数字とが含まれているときは、名刺所有者の勤務先住所とみなすことができる。同様に、「〒」「電話」「TEL」「FAX」などの記号や文字に続いて、特定の桁数の数字の文字列があるときは、その文字列はそれぞれ郵便番号や電話番号とみなすことができる。このような名刺特有の文字や文字列の特徴に基づいて、文字列編集部23は認識文字の属性を決定し、所定の配列で文字列を整理する。なお、本実施例では、4桁の数字で構成された文字列を検査文字4とすることができる。

【0032】上記の如く整理された文字列データは、データ伝送部24によってデータ処理装置6に送り出される。このとき名刺の画像データも文字認識装置5からデータ処理装置6に送り出される。データ処理装置6では、文字認識装置5から伝送された検査文字4を含む文字列データに基づいて、文字列検査部32によって文字列の誤認識の有無を検査する。検査の結果、誤認識がなかったときは、前記文字列データをデータ記憶部34に記憶される。もし、誤認識があることが判明したときは、文字認識装置5から伝送された画像データと、文字列データを図5に示したディスプレイ62に表示し、目視で比較照合し、誤認識を発生した文字列を探し修正する。もし必要なら、再度文字列検査を行い、誤認識を解消したことを精査することもできる。文字列データの修正を完了したらデータ記憶部34に文字列を記録する。

【0033】上記操作を全ての名刺について実行することにより、もし文字列の誤認識が全くなければ人手を介することなく、データ処理装置6のデータ記憶部34に全ての文字列データが格納されていることになる。また、誤認識があったときは、文字列修正部33を操作して文字列を修正のうえ、データ記憶部34に正しい文字列データを格納することができる。なお、データ処理装置6のデータ記憶部34に記録された文字列データは、住所録や顧客管理のプログラムを起動して、それぞれ必要な文字列データを取り込み、宛て名印刷やダイレクトメールの発送に広く応用することができる。

【0034】次に、本発明の検査文字付き印刷物の他の実施例について説明する。図1に示した実施例において

は、印刷物ごとに検査文字を1組だけ使用したが、より高い検査精度を得るために、例えば、図4に示すように1行ごとに検査文字を設け、1行単位で認識文字の検査をしてもよい。図4は、横型名刺の一例である。図において、1は、名刺であり、2は、機械読取する印刷文字、3は、所属団体のロゴマークである。名刺1には、機械読取する印刷文字2の各文字列の右側にそれぞれ4桁の数字で構成された検査文字を配置してある。図4の例では印刷文字列が8行あり、検査文字は各文字列に対応して8行分設けてある。各行の検査文字は、それぞれ対応する印刷文字の行ごとに設定され、従って、文字読取装置で文字認識したときの、誤認識の判別は行単位で行われることになる。

【0035】図に示すように、各行単位に検査文字を付加することによって、万一、文字認識装置で文字を誤認識した場合、データ処理装置における文字列修正を行単位に行うことになり、図1の例と比較して修正作業が容易になる。すなわち、図1の例では全印刷文字に対して1つの検査文字を付加したので、データ処理装置での文字列検査部において誤認識があると判別されても、どの行のどの文字を誤認識しているか容易に見つけることができないが、図4に示す例では、少なくともどの行に誤認識の文字があるかは即座に判定することができる。

【0036】次に、検査文字自体の他の表記例について説明する。以下は「ABC株式会社」を、JISコードと呼ばれるコード体系で16進数で表記したものである。

「A」=2341

「B」=2342

「C」=2343

「株」=3374

「式」=3C30

「会」=3271

「社」=3C52

【0037】前述と同様に、先ず、上記で与えられた、各文字コードの総和を求める。この場合、16進法で総和を求め「1482D」を得る。次いで、前記総和の下4桁「482D」を得る。このようにして得られた4桁の16進数値を文字に変換して検査文字とする。従って、検査文字をJISコードで表記するときは、「0」～「9」の10個の数字と「A」～「F」の6個の英字を用いて表記することになる。

【0038】上記16進数の英数字による検査文字の場合、文字列検査部において、誤認識の有無を検査するとき、文字認識装置で認識した文字列データの各文字の文字コードの総和を求め、次いで検査文字と比較し、一致するか否かにより誤認識の有無を検査する。すなわち、文字コードの総和から得られた下4桁の16進数の数値と検査文字から得られた検査値とを比較し、一致したときは誤認識がなかったと判断し、一致しなかったときは

誤認識があったと判断することができる。

【0039】なお、名刺に使用される特殊な記号として、郵便番号を表す「〒」や、電話番号を表す電話機のロゴマークも、文字として認識することにより、より高精度の文字認識が可能になる。また、本実施例では、名刺における文字認識について説明したが、他の印刷物、例えば雑誌や単行本などの書籍や、新聞などにおいても使用できることは明かである。

【0040】さらに、本実施例では4桁（4文字）の検査文字を使用した。文字認識装置の誤認識の確率が低いときは、より少ない桁数の検査文字を使用してもよい。

【0041】

【発明の効果】以上詳細に説明した如く本発明の印刷物は、請求項1に示すように、漢字を含む日本語の機械読取対象文字列に対して、検査文字を付したので、文字認識装置にて読み取った文字列における、文字認識ミスの有無を機械的に検査することができ、文字列データの検証作業を省力化できる。

【0042】また、本発明の印刷物は、請求項2に示すように、請求項1の発明に加えて、名刺であるので、名刺読取装置で読み取った名刺情報の、認識ミスを簡便に検査することができる。

【0043】さらに、本発明の印刷物は、請求項3に示すように、請求項1または請求項2の発明に加えて、検査文字が10進数数字で構成されているので、名刺のデ

ザインを損なうことなく、また、検査文字自体の認識間違いが少なく、かつ、文字認識ミスに対する高い検査精度が得られ、信頼性の高い文字検査ができる。

【0044】また、本発明の印刷物は、請求項4に示すように、請求項1または請求項2の発明に加えて、検査文字が16進数英数字で構成されているので、より高い文字認識ミスの検査精度が得られる。

【図面の簡単な説明】

【図1】 本発明の印刷物の一実施例を示す平面図である。

【図2】 本発明の印刷物の文字認識結果を示すデータ例である。

【図3】 文字読取装置の構成例を示すブロック図である。

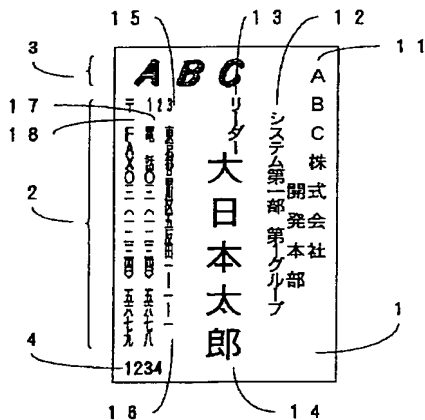
【図4】 本発明の印刷物の他の実施例を示す平面図である。

【図5】 文字読取装置の一例を示す外観斜視図である。

【符号の説明】

- 1 名刺
- 2 印刷文字
- 3 ロゴマーク
- 4 検査文字
- 5 文字認識装置
- 6 データ処理装置

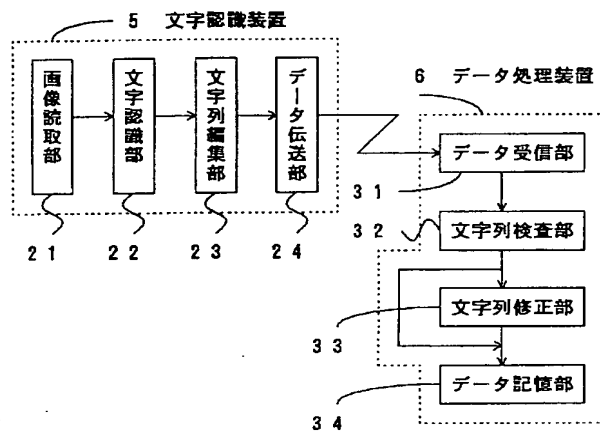
【図1】



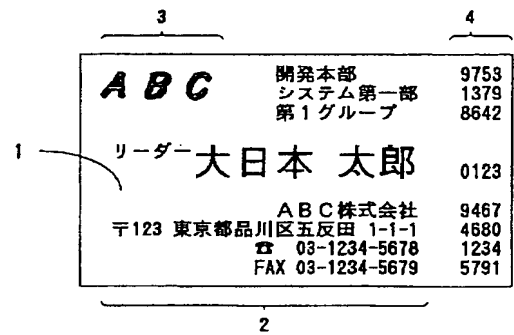
【図2】

11	団体名	ABC株式会社
12	所 属 部 署	1 開発本部 2 システム第一部 3 第1グループ
13	役 職 名	リーダー
14	姓 名	大日本太郎
15	所 在 地	東京都品川区五反田1丁目1番1号
16	郵便番号	123
17	電話番号	03-1234-5678
18	FAX番号	03-1234-5679
4	検査文字	1234

【図3】



【図4】



【図5】

